



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/644,815	08/21/2003	Jerome R. Bellegarda	P2989-908	6190
62665	7590	05/12/2010	EXAMINER	
BUCHANAN INGERSOLL & ROONEY, PC			DWIVEDI, MAHESH H	
P.O. BOX 1404			ART UNIT	PAPER NUMBER
ALEXANDRIA, VA 22313-1404			2168	
NOTIFICATION DATE	DELIVERY MODE			
05/12/2010	ELECTRONIC			

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ADIPFDD@bipc.com  
offserv@bipc.com

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	10/644,815	BELLEGARDA ET AL.	
	<b>Examiner</b>	<b>Art Unit</b>	
	MAHESH H. DWIVEDI	2168	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

1) Responsive to communication(s) filed on 22 April 2010.  
 2a) This action is **FINAL**.                            2b) This action is non-final.  
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

4) Claim(s) 1-7,9-11,13-23,25-28,30-33,35-38 and 48-58 is/are pending in the application.  
 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.  
 5) Claim(s) \_\_\_\_\_ is/are allowed.  
 6) Claim(s) 1-7,9-11,13-23,25-28,30-33,35-38 and 48-58 is/are rejected.  
 7) Claim(s) \_\_\_\_\_ is/are objected to.  
 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

9) The specification is objected to by the Examiner.  
 10) The drawing(s) filed on 21 August 2003 is/are: a) accepted or b) objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).  
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
 a) All    b) Some \* c) None of:  
 1. Certified copies of the priority documents have been received.  
 2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)	4) <input type="checkbox"/> Interview Summary (PTO-413)
2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)	Paper No(s)/Mail Date. _____ .
3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)	5) <input type="checkbox"/> Notice of Informal Patent Application
Paper No(s)/Mail Date _____.	6) <input type="checkbox"/> Other: _____ .

## DETAILED ACTION

### ***Continued Examination Under 37 CFR 1.114***

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 4/22/2010 has been entered.

### ***Remarks***

2. Receipt of Applicant's Amendment, filed on 04/22/2010, is acknowledged. The amendment includes the cancellation of claims 8, 12, 24, 29, 34, and 39-47, and the addition of claims 58.

### ***Claim Rejections - 35 USC § 103***

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, 48, 52, 54, 56, and 58 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Bellegarda et al.** (Article entitled "Exploiting Latent Semantic Information in Statistical Language Modeling", dated 10/26/2000) and in view of **Vivisimo** (Article entitled "Vivisimo FAQ", dated 02/04/2002).

5. Regarding claim 1, **Bellegarda** teaches a method comprising:

- A) mapping the files into a semantic vector space (Page 1279, Abstract);
- B) clustering the files within said space (Page 1279, Abstract).
- C) wherein multiple threshold values that are settable to desired levels of granularity are defined, and said files are clustered based on said multiple threshold values (Page 1284)

The examiner notes that **Bellegarda** teaches “**mapping the files into a semantic vector space**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**clustering the files within said space**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**wherein multiple threshold values that are settable to desired levels of granularity are defined, and said files are clustered based on said multiple threshold values**” as “Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space” (1284).

**Bellegarda** does not explicitly teach:

- D) deriving a hierarchy of plural level of clusters from said clustering;
- E) displaying the files in a hierarchical format of plural level of clusters based on said derived hierarchy.

**Vivisomo**, however, teaches “**deriving a hierarchy of plural level of clusters from said clustering**” as “Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there

is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísmo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04), and **"displaying the files in a hierarchical format of plural level of clusters based on said derived hierarchy"** as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísmo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and

"soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 2, **Bellegarda** does not explicitly teach a method comprising:  
A) wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system.

**Vivisimo**, however, teaches "**wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system**" as "Clustering is done just before the user sees the search results, just in time. There is no need to prepare anything beforehand, much less pre-process the entire document collection from where the results came. Clustering is a fully automatic process that requires no preparation steps, and hence no maintenance. Classification requires pre-specifying categories (typically broad and hence rather bland) and updating these categories as new documents are added to the collection" (Page 03).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 3, **Bellegarda** further teaches a method comprising:

A) wherein the step of clustering the files is performed in response to the creation of a new file within the file system (Page 1286, Section: A. Framework Extension).

The examiner notes that **Bellegarda** teaches “**wherein the step of clustering the files is performed in response to the creation of a new file within the file system**” as “finding a new representation for a new document in the space S is straightforward” (Page 1286, Section: A. Framework Extension). The examiner further notes that it is clear that the method of **Bellegarda** clusters when a new document is noticed.

Regarding claim 4, **Bellegarda** further teaches a method comprising:

A) wherein said files are text documents (Page 1279, Abstract); and  
B) said mapping is conducted on the basis of a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches “**wherein said files are text documents**” as “This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**said mapping is conducted on the basis of a language model**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract).

Regarding claim 5, **Bellegarda** further teaches a method comprising:

A) wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and  
B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector**” as “The starting point is the construction of a matrix ( $W$ ) of co-occurrences between words and documents” (Page 1281, Section: A. Feature Extraction) and “The ( $M \times N$ ) word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$  (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches “**associates each document with a vector**” as “The ( $M \times N$ ) word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$ ” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 6, **Bellegarda** further teaches a method comprising:  
A) the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space**” as “To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 7, **Bellegarda** further teaches a method comprising:  
A) wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches “**wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another**” as “This opens up the opportunity to apply familiar clustering techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space” (Page 1286, Section: A. Framework Extension).

Regarding claim 9, **Bellegarda** does not explicitly teach a method comprising:  
A) including the step of automatically labeling the clusters based on the resulting clusters.

**Vivisimo**, however teaches “**including the step of automatically labeling the clusters based on the resulting clusters**” as “Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case” (Page 03) and “Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results” (Page 03), “Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results” (Page 03), and “We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human

being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 10, **Bellegarda** does not explicitly teach a method comprising:  
A) wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

**Vivisimo**, however teaches "**wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03) and "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely

mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), and "We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 11, **Bellegarda** teaches a graphical user interface comprising:

- A) a virtual file system (Page 1279, Abstract); and
- B) clustering said files based on multiple threshold values that are settable to desired levels of granularity (Page 1284).

The examiner notes that **Bellegarda** teaches "**a virtual file system with a semantic hierarchy, wherein the semantic hierarchy is based on clustering of files based on semantic similarities**" as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches "**clustering said files based on multiple threshold values that are settable to desired levels of granularity**" as "Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain

a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

**Bellegarda** does not explicitly teach:

- A) A graphical user interface configured to display files with a semantic hierarchy of plural levels of clusters that is derived from semantic similarities of said files;
- C) determining a directory structure having plural levels of clusters based on the clustering determined from similarities between said files, wherein the graphical user interface graphically presents the determined directory structure having plural levels of clusters to be displayed on a display device.

**Vivisimo**, however, teaches "**A graphical user interface configured to display files with a semantic hierarchy of plural levels of clusters that is derived from semantic similarities of said files**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivisimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this

is better than forcing documents to fit in a single location" (Page 04), and "**determining a directory structure having plural levels of clusters based on the clustering determined from similarities between said files, wherein the graphical user interface graphically presents the determined directory structure having plural levels of clusters to be displayed on a display device**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 13, **Bellegarda** does not explicitly teach a graphical user interface comprising:

- A) wherein clustering of the files is initiated by user selection.

**Vivisimo**, however, teaches “**wherein clustering of the files is initiated by user selection**” as “Clustering is done just before the user sees the search results, just in time. There is no need to prepare anything beforehand, much less pre-process the entire document collection from where the results came. Clustering is a fully automatic process that requires no preparation steps, and hence no maintenance. Classification requires pre-specifying categories (typically broad and hence rather bland) and updating these categories as new documents are added to the collection” (Page 03).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo’s** would have allowed **Bellegarda’s** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 14, **Bellegarda** further teaches a graphical user interface comprising:

- A) wherein clustering of the files is initiated upon creation of a new file in the file system (Page 1286, Section: A. Framework Extension).

The examiner notes that **Bellegarda** teaches “**wherein clustering of the files is initiated upon creation of a new file in the file system**” as “finding a new representation for a new document in the space S is straightforward” (Page 1286, Section: A. Framework Extension). The examiner further notes that it is clear that the method of **Bellegarda** clusters when a new document is noticed.

Regarding claim 15, **Bellegarda** further teaches a graphical user interface comprising:

- A) wherein text files are clustered utilizing a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches “**analyzing files in a file system to determine similarities in data pertaining to their content**” as “(discrete) words

and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

**Bellegarda** does not explicitly teach:

B) non-text files are clustered utilizing rule-based techniques.

**Oliver**, however, teaches "**non-text files are clustered utilizing rule-based techniques**" as "Vivísimo now also supports the most advanced features of the major search engines using one Vivísimo syntax, which follows the most standard conventions. Vivísimo translates your query into the corresponding syntax of each underlying search engine. Vivísimo only queries the search engines that support your chosen syntax. (Check which engines have been queried by clicking on the Details link at the top of the results page.) Thus, you can safely use +,-,... or common Boolean operators (**NEAR,OR,...**) as well as common field searches such as **image:, title:, link:...** or search restrictions such as **host: or domain:...**" (Page 08).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivísimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivísimo** (Page 04).

Regarding claim 16, **Bellegarda** further teaches a graphical user interface comprising:

A) wherein said language model comprises the LSA paradigm (Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches "**wherein said language model comprises the LSA paradigm**" as "The focus of this paper is on semantically driven span extension only, and more specifically on how the LSA paradigm can be exploited to improve statistical language modeling" (Page 1281, Section: D. Organization).

Regarding claim 17, **Bellegarda** teaches a computer-readable media comprising:

- A) analyzing files in a file system to determine similarities in data pertaining to their content (Page 1279, Abstract);
- B) clustering said files based on multiple threshold values that are settable to desired levels of granularity (1284);

The examiner notes that **Bellegarda** teaches “**analyzing files in a file system to determine similarities in data pertaining to their content**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**clustering said files based on multiple threshold values that are settable to desired levels of granularity**” as “Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space” (1284).

**Bellegarda** does not explicitly teach:

- C) determining a directory structure having plural levels of clusters based on the clustering determined from similarities between the files;
- D) displaying files in hierarchical format of plural levels of clusters based on the clustering determined from similarities between the files.

**Vivisimo**, however, teaches “**determining a directory structure having plural levels of clusters based on the clustering determined from similarities between the files**” as “Document clustering is the automatic organization of documents into

groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísmo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04), and **"displaying files in hierarchical format of plural levels of clusters based on the clustering determined from similarities between the files"** as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísmo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping

an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 18, **Bellegarda** further teaches a computer-readable media comprising:

- A) wherein said files are text documents (Page 1279, Abstract); and
- B) the similarities are based upon the word content of the files (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches "**wherein said files are text documents**" as "This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches "**the similarities are based upon the word content of the files**" as "The starting point is the construction of a matrix ( $W$ ) of co-occurrences between words and documents" (Page 1281, Section: A. Feature Extraction) and "The ( $M \times N$ ) word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely

associated with a column vector of dimension  $M$  (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 19, **Bellegarda** further teaches a computer-readable media comprising:

- A) wherein said similarities are determined in accordance with a language model (Page 1279, Abstract, Page 1281, Section: D. Organization); and
- B) the files are clustered in accordance with said model (Page 1279, Abstract, Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches “**wherein said similarities are determined in accordance with a language model**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**the files are clustered in accordance with said model**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract).

Regarding claim 20, **Bellegarda** further teaches a computer-readable media comprising:

- A) wherein said language model comprises the LSA paradigm (Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches “**wherein said language model comprises the LSA paradigm**” as “The focus of this paper is on semantically driven span extension only, and more specifically on how the LSA paradigm can be exploited to improve statistical language modeling” (Page 1281, Section: D. Organization).

Regarding claim 21, **Bellegarda** further teaches a computer-readable media comprising:

- A) wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and
- B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector**” as “The starting point is the construction of a matrix ( $W$ ) of co-occurrences between words and documents” (Page 1281, Section: A. Feature Extraction) and “The ( $M \times N$ ) word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$  (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches “**associates each document with a vector**” as “The ( $M \times N$ ) word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$ ” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 22, **Bellegarda** further teaches a computer-readable media comprising:

- A) wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space

(Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space**” as “To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 23, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches “**wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another**” as “This opens up the opportunity to apply familiar clustering techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space” (Page 1286, Section: A. Framework Extension).

Regarding claim 25, **Bellegarda** does not explicitly teach a computer-readable media comprising:

A) wherein said computer-executable code performs step of automatically labeling the clusters based on the resulting clusters.

**Vivisimo**, however teaches “**wherein said computer-executable code performs step of automatically labeling the clusters based on the resulting clusters**” as “Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human

intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03) and "Instead of producing a flat list of groups, Vivísmo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísmo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), and "We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 26, **Bellegarda** does not explicitly teach a computer-readable media comprising:

- A) wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

**Vivisimo**, however teaches “**wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster**” as “Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case” (Page 03) and “Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results” (Page 03), “Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results” (Page 03), and “We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up” (Page 04)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 27, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein the computer executable code performs the following steps: clustering text files within the file system using semantic similarities (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches “**a semantic hierarchy that is based upon the content of said files**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract).

**Bellegarda** does not explicitly teach:

- B) clustering non-text files within the files system using rule-based techniques;
- C) labeling the resulting clusters; and
- D) displaying the files in a hierarchical format based on the resulting clusters and labels.

**Olliver**, however, teaches “**clustering non-text files within the files system using rule-based techniques**” as “Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case” (Page 03) and “Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results” (Page 03), and “Conceptual clustering

methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), "**labeling the resulting clusters**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03) and "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "**Conceptual clustering** methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), and "We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04), and "**displaying the files in a hierarchical format based on the resulting clusters and labels**" as "Document clustering is the automatic

organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivisimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 28, **Bellegarda** teaches a computer system comprising:

- A) a file system storing files (Page 1279, Abstract);
- C) a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space, cluster the files within said space based on multiple

threshold values that are settable to desired levels of granularity (Pages 1279 and 1284, Abstract);

The examiner notes that **Bellegarda** teaches “**a file system storing files**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space, cluster the files within said space based on multiple threshold values that are settable to desired levels of granularity**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract) and “Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space” (1284).

**Bellegarda** does not explicitly teach:

- B) a display device; and
- D) derive a hierarchy of plural levels of clusters from said clustering;
- E) a user interface which displays representations of files stored in said file system in the form of said derived hierarchy of plural level of clusters.

**Vivisimo**, however, teaches “**a display device**” as “Document clustering is the automatic organization of documents into groups or clusters. “Document clustering” differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case” (Page 03), “Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known “Windows Explorer”-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results” (Page 03), “No. Simple one-word queries often lead to clusters that modify the query. For example, “soap” can lead to “soap opera”, “handmade soap”, and “soap bubbles”, but also to “simple object access protocol”, known also by its SOAP acronym” (Page 04), and “No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location” (Page 04), “**derive a hierarchy of plural levels of clusters from said clustering**” as “Document clustering is the automatic organization of documents into groups or clusters. “Document clustering” differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case” (Page 03), “Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known “Windows Explorer”-style interface. This interface can be used with no training since it is quite

intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04), and "**a user interface which displays representations of files stored in said file system in the form of said derived hierarchy of plural level of clusters**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivisimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 30, **Bellegarda** further teaches a computer system comprising:

- A) wherein said files are text documents (Page 1279, Abstract); and
- B) said processor maps said files on the basis of a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches “**wherein said files are text documents**” as “This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus” (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches “**said processor maps said files on the basis of a language model**” as “(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties” (Page 1279, Abstract).

Regarding claim 31, **Bellegarda** further teaches a computer system comprising:

- A) wherein said processor constructs a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and
- B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**wherein said processor constructs a matrix which associates each word in the documents with a vector**” as “The starting point is the construction of a matrix ( $W$ ) of co-occurrences between words and documents” (Page 1281, Section: A. Feature Extraction) and “The ( $M \times N$ )

word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$  (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches “**associates each document with a vector**” as “The  $(M \times N)$  word-document matrix  $W$  resulting from the above feature extraction defines two vector representations for the words and the documents. Each word  $\omega_i$  can be uniquely associated with a row vector of dimension  $N$ , and each document  $d_j$  can be uniquely associated with a column vector of dimension  $M$ ” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 32, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches “**wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space**” as “To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis” (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 33, **Bellegarda** further teaches a computer system comprising:

A) wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches “**wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another**” as “This opens up the opportunity to apply familiar clustering

techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space" (Page 1286, Section: A. Framework Extension).

Regarding claim 35, **Bellegarda** does not explicitly teach a computer system comprising:

A) wherein said processor automatically labels the clusters based on the resulting clusters.

**Vivisimo**, however teaches "**wherein said processor automatically labels the clusters based on the resulting clusters**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03) and "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), and "We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not

perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 36, **Bellegarda** does not explicitly teach a computer system comprising:

A) wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

**Vivisimo**, however teaches "**wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03) and "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "Conceptual clustering methods interleave the process of forming groups with the step of annotating them, much like people might do by hand. So, if Vivísimo tries to form a group but judges that the group cannot be described well, the group is rejected. In contrast, some other approaches rely mainly on mathematical optimization, in which description of the groups

is relegated to the end after the groups are formed, which gives generally worse results" (Page 03), and "We are gratified that users sometimes ask this. The annotations are created spontaneously by the software. When they are good, it seems that a human being must have created the categories and the machine merely recognizes the documents that belong there, which is not the case. However, our technology is not perfect: the diligent user will surely spot an occasional annotation that only a machine would make up" (Page 04)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 37, **Bellegarda** does not explicitly teach a method comprising:  
A) wherein said deriving step includes organizing the clusters into a hierarchical directory structure.

**Vivisimo**, however, teaches "**wherein said deriving step includes organizing the clusters into a hierarchical directory structure**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap

opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 38, **Bellegarda** teaches a method comprising:

- A) mapping all words of the plurality of documents and the plurality of documents in a semantic vector space (Page 1279, Abstract);
- B) generating a plurality of clusters based on the semantic similarities of the plurality of documents and multiple threshold values that are settable to desired levels of granularity (Pages 1279 and 1284, Abstract).

The examiner notes that **Bellegarda** teaches "**mapping all words of the plurality of documents and the plurality of documents in a semantic vector space**" as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches "**generating a plurality of clusters based on the semantic similarities of the plurality of documents and multiple threshold values that are settable to desired levels of granularity**" as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic

classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract) and "Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

**Bellegarda** does not explicitly teach:

- C) organizing the plurality of clusters into directories in a hierarchical format of plural levels of clusters;
- D) displaying the plurality of documents in said hierarchical format of plural levels of clusters based on a result of clustering the plurality of documents.

**Vivisimo**, however, teaches "**organizing the plurality of clusters into directories in a hierarchical format of plural levels of clusters**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that

modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04), and "**displaying the plurality of documents in said hierarchical format of plural levels of clusters based on a result of clustering the plurality of documents**" as "Document clustering is the automatic organization of documents into groups or clusters. "Document clustering" differs from other techniques (classification, taxonomy building, Northern Light, etc.) in that it is fully automated: there is no human intervention at any point (except that people wrote the basic algorithms). The biggest challenge for document clustering has been to quickly find meaningful groups that are concisely annotated. Our innovation relies on a newly discovered heuristic algorithm that does this well. Our clustering algorithm has achieved good results on web pages, patent abstracts, newswires, meeting transcripts, and television transcripts with little or no customization in every case" (Page 03), "Instead of producing a flat list of groups, Vivísimo organizes groups into a hierarchy or tree, using a well-known "Windows Explorer"-style interface. This interface can be used with no training since it is quite intuitive. Users can zoom in on items of interest while keeping an overview of all the search results" (Page 03), "No. Simple one-word queries often lead to clusters that modify the query. For example, "soap" can lead to "soap opera", "handmade soap", and "soap bubbles", but also to "simple object access protocol", known also by its SOAP acronym" (Page 04), and "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

The examiner further notes that the non-applied art of **Arnold** shows an interface of the **Vivisimo** search engine. Specifically, there is shown clusters of hierarchical folders that allow a user to drill down further if need be.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching

**Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

Regarding claim 48, **Bellegarda** further teaches a method comprising:

A) wherein the multiple threshold values are characteristic values of clusters from said clustering (Page 1284).

The examiner notes that **Bellegarda** teaches “**wherein the multiple threshold values are characteristic values of clusters from said clustering**” as “Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$  . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space” (1284).

Regarding claim 50, **Bellegarda** further teaches a graphical user interface comprising:

A) wherein the multiple threshold values are characteristic values of clusters from said clustering (Page 1284).

The examiner notes that **Bellegarda** teaches “**wherein the multiple threshold values are characteristic values of clusters from said clustering**” as “Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final

set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

Regarding claim 52, **Bellegarda** further teaches a computer readable media comprising:

A) wherein the multiple threshold values are characteristic values of clusters from said clustering (Page 1284).

The examiner notes that **Bellegarda** teaches "wherein the multiple threshold values are characteristic values of clusters from said clustering" as "Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

Regarding claim 54, **Bellegarda** further teaches a computer system comprising:  
A) wherein the multiple threshold values are characteristic values of clusters from said clustering (Page 1284).

The examiner notes that **Bellegarda** teaches "**wherein the multiple threshold values are characteristic values of clusters from said clustering**" as "Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final

set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

Regarding claim 56, **Bellegarda** further teaches a method comprising:

A) wherein the multiple threshold values are characteristic values of clusters from said clustering (Page 1284).

The examiner notes that **Bellegarda** teaches "**wherein the multiple threshold values are characteristic values of clusters from said clustering**" as "Once (11) is specified, it is straightforward to proceed with the clustering of the word vectors , using any of a variety of algorithms (see, for instance, [2]). Since the number of such vectors is relatively large, it is advisable to perform this clustering in stages, using, for example, K-means and bottom-up clustering sequentially. In that case, K-means clustering is used to obtain a coarse partition of the vocabulary in to a small set of superclusters. Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters  $C_k$ ,  $1 \leq k \leq K$ , . This process can be thought of as uncovering, in a data-driven fashion, a particular layer of semantic knowledge in the space" (1284).

Regarding claim 58, **Bellegarda** does not explicitly teach a method comprising:  
A) providing a user an option to reorganize the files in the file system according to the derived hierarchy.

**Vivisimo**, however, teaches "**providing a user an option to reorganize the files in the file system according to the derived hierarchy**" as "No. Sometimes a document fits well in more than one place in the hierarchy, so we place it everywhere it fits. For users, this is better than forcing documents to fit in a single location" (Page 04).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Vivisimo's** would have allowed **Bellegarda's** to provide a clustering that is user friendly, concise, and fast, as noted by **Vivisimo** (Page 04).

6. Claims 49, 51, 53, 55, and 57 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Bellegarda et al.** (Article entitled "Exploiting Latent Semantic

Information in Statistical Language Modeling, dated 10/26/2000) and in view of **Vivisimo** (Article entitled "Vivisimo FAQ", dated 02/04/2002) as applied to claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, 48, 52, 54, 56, and 58 above, and further in view of **Hertz** (U.S. PGPUB 2003/0037041).

7. Regarding claims, 49, 51, 53, 55, and 57, **Bellegarda** and **Oliver** do not explicitly teach a method, graphical user interface, computer-readable media, computer system, and computer comprising:

A) wherein the characteristic values of the clusters are cluster variances of the clusters.

**Hertz**, however, teaches "**wherein the characteristic values of the clusters are cluster variances of the clusters**" as "a real number determined by calculating the statistical variance of the profiles of all target objects in a cluster, is termed a "cluster variance,"" (Paragraph 13) and "The threshold used in step 6 is typically an affine function or other function of the greater of the cluster variances (or cluster diameters) of S and T" (Paragraph 326).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Hertz's** would have allowed **Bellegarda's** and **Oliver's** to provide for a more efficient method in gathering data that interests users, as noted by **Hertz** (Paragraph 11).

#### ***Response to Arguments***

8. Applicant's arguments with respect to claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58, have been considered but are moot in view of the new ground(s) of rejection.

#### ***Conclusion***

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

U.S. Patent 6,820,094 issued to **Ferguson et al.** on 16 November 2004. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. PGPUB 2004/0249865 issued to **Lee et al.** on 09 December 2004. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to automatically name and label folders).

U.S. PGPUB 2004/0148453 issued to **Watanabe et al.** on 29 July 2004. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to automatically name and label folders).

U.S. Patent 5,819,258 issued to **Vaithyanathan et al.** on 06 October 1998. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. Patent 6,360,227 issued to **Aggarwal et al.** on 19 March 2002. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. Patent 5,899,995 issued to **Millier et al.** on 04 May 1999. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. Patent 7,158,986 issued to **Oliver et al.** on 02 January 2007. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. Patent 7,085,767 issued to **Kusama** on 01 August 2006. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

Article entitled “Vivisimo: Clustering Delivers Information Overlook”, dated 05/03/2003, by **Arnold**. The subject matter disclosed therein is pertinent to that of claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, and 48- 58 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

***Contact Information***

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Mahesh Dwivedi whose telephone number is (571) 272-2731. The examiner can normally be reached on Monday to Friday 8:20 am – 4:40 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Tim Vo can be reached (571) 272-3642. The fax number for the organization where this application or proceeding is assigned is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Mahesh Dwivedi  
Patent Examiner  
Art Unit 2168

May 09, 2010  
/Mahesh H Dwivedi/  
Examiner, Art Unit 2168